
WORKER PROFILING AND REEMPLOYMENT SERVICES
Profiling Methods: Lessons Learned

by Marisa Kelso
June, 1998

TABLE OF CONTENTS

I. INTRODUCTION	3
II. PROFILING METHODS	3
A. Overview	3
(i) The Department of Labor Model	4
(ii) State Versions of the DOL Model	6
(iii) State Models Beyond the DOL Model	6
B. The Dependent Variable	7
(i) Alternative Specifications of the Dependent Variable	8
(ii) Adjusting the Coding Scheme of the Exhaustion Variable	9
C. Independent Variables	11
(i) DOL Model Core Variables	11
(ii) Data Elements Beyond the DOL Model	14
(iii) Addressing Sub-state Labor Markets	18
(iv) Developing Sub-state Models	19
III. POLICY IMPLICATIONS OF PROFILING METHODS	20
IV. CONTINUING CONCERNS	21
APPENDIX	22

WORKER PROFILING AND REEMPLOYMENT SERVICES

Profiling Methods: Lessons Learned

I. INTRODUCTION

It has been five years since the Unemployment Compensation Amendments of 1993 spawned the Worker Profiling and Reemployment Services (WPRS) system. In that time, we have laid the foundation, designed its architecture, and provided guidance to those who forged its structure. We have encountered many obstacles on the road to completion, but the resulting experiences have lead us to the discovery of useful insights about existing organizations; including their methodologies of choice, procedures and the trials of inter-agency cooperation. Since its inception, the “worker profiling program [has encouraged] States to use the Unemployment Insurance system to link permanently displaced workers to reemployment services early in their period of unemployment and facilitate their transition to new jobs”¹ The concept of “linking” and joining efforts to enhance the results of the program was in the foreground of the project, as early as President Clinton’s statement on the day he signed the original bill into law. This paper continues along the initial theme of collaboration, emphasizing the symbiosis that takes place in a cooperative environment. Technical methodologies and inter-organizational linkages have taken many forms during the WPRS system implementation, but throughout each variation, a common thread remains: cooperation is the key to the success of both this program, and the future o four organizations.

This premise of cooperation has taken form in our practical application of the WPRS system. A pattern throughout the implementation process has been the recurring objective of states to draw on other states’ experiences in establishing and refining WPRS systems. As a primary step toward providing a forum for cooperation, staff from 13 states with whom the TA team has had occasion to work or to contact were canvassed. Table A (Appendix) lists these states and the model specifications they contributed. This is not meant to be an exhaustive survey of all the methods that have been tested thus far or that may prove effective in the future. Rather, the discussion summarizes techniques used by states to identify data elements historically correlated with benefit exhaustion and to incorporate these elements into a chronicle of profiling development. Based on the experiences of the TAT, it provides a broad, nationwide perspective on the lessons learned throughout the state implementation process, focusing on successful strategies and hopefully providing a basis for a continued exchange of similar information.

¹Statement by the President, Office of the Press Secretary. March 4, 1993. <<http://www.whitehouse.gov>>, Virtual Library.

II. PROFILING METHODS

A. Overview

WPRS attempts to identify unemployment insurance (UI) claimants with a high potential for exhausting their benefits and provide them with re-employment services. These claimants represent a demand for services, with new and existing programs representing the supply. Prior to WPRS, the demand for and supply of re-employment services were not necessarily balanced. WPRS is a tool which facilitates both the identification of claimants and the allocation of services, such that those claimants most likely to exhaust receive highest priority in receiving available re-employment services. This work provides a sampling of state experiences thus far in designing identification methods for the WPRS initiative and should provide useful ideas for states as they continue to develop and refine their profiling mechanisms. It is important for readers to note that this work is informational only, and does not imply that inclusion of these new variables or approaches is necessary for a successful WPRS experience.

To make the necessary identifications, states may use either characteristic screens or a statistical model. Both methods seek to identify characteristics common to recent exhaustees and target current claimants who share these characteristics. Although neither method can target exhaustees with complete accuracy, both screens and models have been found considerably more accurate than less-systematic and less-scientific processes such as random selection. Most states have chosen to implement statistical models, since they offer both greater accuracy and greater procedural flexibility than do characteristic screens. A few states without sufficient historical data to develop a statistical model have chosen to implement screening methodologies and have taken steps to collect data necessary to develop a model in the future. Most of the concepts noted in this paper apply to statistical modelling since it is the more complex and widely used of the two. Many of the strategies and data elements mentioned could be incorporated into a screening methodology as well.

With either method, the target population of WPRS as specified in P.L. 103-152 is claimants who are "likely to exhaust." While the specific make-up of this population changes from state to state, the ultimate goal is to identify claimants whose job-search skills are no longer sufficient to obtain suitable employment in their particular labor market. Identifying these potential exhaustees, while theoretically straightforward, becomes complicated in the practical application of WPRS for a number of reasons. For example, the availability and integrity of historical data are issues in many states. Data from separate intake systems must often be merged, causing additional problems. And some readily available data elements depicting personal characteristics (e.g., ethnicity) have been determined to be discriminatory under Federal equal opportunity legislation and are thus prohibited. Finally, some key influences on benefit exhaustion, such as motivation and networking skills, are not quantifiable; these would affect whether or not a claimant will exhaust his/her benefits but can neither be captured nor factored into a model. Given these complicating influences, the theoretically straightforward problem becomes more difficult to unravel. To which the experiences of many states attest, it is a

formidable task to develop an identification mechanism for WPRS that can accurately predict which new claimants will become exhaustees.

(i) The Department of Labor Model

Rising to these challenges, states have moved forward with profiling. Although predicting exhaustion is an inexact science, states have been able to develop models that considerably reduce prediction errors relative to less-rigorous methods. As mentioned, most have either directly adopted the model initially developed by DOL in 1993, or used it as a starting point in developing a state-specific strategy for identifying likely exhaustees.² The model consists of two initial screens--recall status and union hiring hall--and a set of variables derived from five data elements--education, job tenure, industry, occupation and local unemployment rate. Originally developed from a national data set, the DOL model was first adapted to state-level data in the test state of Maryland. The national and Maryland versions are compared in the following table:

²The DOL model was initially outlined in UI Information Bulletin 4-94 and Field Memorandum 35-94 and was updated in UI Information Bulletins 11-94 and 15-94. These and all WPRS-related issuances can be found in Unemployment Insurance Occasional Paper 94-4, "The Worker Profiling and Reemployment Services System: Legislation, Implementation Process and Research Findings."

Table 1. National and Maryland Model Comparisons

	NATIONAL MODEL	MARYLAND MODEL
EDUCATION	Categorical variables: -Less than HS diploma -HS diploma -Some college -College degree	Categorical variables -Less than HS diploma -HS diploma -Some college -Bachelors degree -Masters degree/PhD
JOB TENURE	Categorical variables: -0-3 years -3-5 years -6-9 years 10+ years	Continuous variable: -Years of job tenure
INDUSTRY	Employment change (%): -SIC Division level -State level	Employment change (%): -SIC Division level -SDA level
OCCUPATION	Binary variable, from employment change (%): -(=1) if growing -(=0) if zero or declining -SOC one-digit level	Categorical variables -DOT one-digit level (nine categories)
UNEMPLOYMENT RATE	Unemployment Rate (%): -State level	Unemployment Rate (%): -SDA level

(ii) State Versions of the DOL Model

Both of the above variations of the DOL model served as starting points in the development of state WPRS identification mechanisms. The national analysis demonstrated on an aggregate level that the five data elements shown above were both logically and statistically correlated with UI benefit exhaustion. The Maryland test state project showed further that constructing a state-specific version of the DOL model would require an additional degree of testing and experimentation. Equally important, the Maryland project demonstrated that an operational state system could be readily developed from the model. This progression is what is meant by the phrase "using a state-specific version of the DOL model" which appears throughout many of the WPRS-related issuances. The same five data elements are included, but depending on an analysis of how (or if) these elements influence exhaustion in the given state, their treatment in the model may differ. As Table 1 shows, the national and Maryland models are different in the way each of the five data elements are treated in the model.

A number of states have followed the Maryland experience closely, using the same set or a very similar set of data elements to construct a simple state-level statistical model. This generally results in a methodology that, when applied to out-of-sample historical data (i.e., data not used to develop the model), is able to correctly identify a higher percentage of claimants as exhaustees compared to the alternatives of random selection and characteristic screening. The Department has contracted with Social Policy Research Associates (SPR) to provide a nationwide analysis and report to congress on follow-up and outcome issues.³ However, for the purposes of this paper, models and data elements are most easily evaluated based upon their performance in analyzing historical data. Section C-(i) summarizes the findings of the states included in this study relative to the five data elements comprising the DOL model.

(iii) State Models Beyond the DOL Model

Alternative State models represent a variation on the basic form and concept of the DOL model: developing state specific processes of identifying and serving likely exhaustees. Incorporating some or all of the core DOL model elements into a statistical model allows states to identify a greater percentage of exhaustees than is possible with other approaches. However, since there exists considerable diversity among states, it is not surprising that several have found that alternative specification are needed to effectively model their populations. And since SESA automated data processing systems retain a great deal more information than just these five elements, several states have expanded upon the DOL model by testing new data elements and variables in an effort to increase predictive ability. States that have done so have at least used the DOL-model elements as a starting point, retaining those found to be helpful in identifying likely exhaustees and building upon them. Other extensions in addition to testing new variables include using alternative statistical methods, and in the case of Kentucky, Washington and Alaska, developing multiple sub-state models.

The following two sections summarize states' experiences developing WPRS models using the DOL-model data elements as a frame of reference. Section B examines issues related to the dependent variable while section C focuses on the independent variables. Within each section, descriptions of data elements and related issues are followed by evaluations of the advantages and/or disadvantages of incorporating each element. These evaluations reflect both the experiences to date of the states included in this study and the assessment of members of the TAT. The intent is to provide worthwhile feedback and direction for states that continue to develop and refine identification methods as WPRS progresses. This feedback should supplement, not substitute for, state-specific analysis of historical data in developing the most practical and effective means of identifying likely exhaustees.

³The findings of these studies are published in the "evaluation of Worker Profiling and Reemployment Services Systems: Interim Report", Unemployment Insurance Occasional Paper 96-1 and "Evaluation of Worker Profiling and Reemployment Services Systems", Report to Congress, 1997.

B. The Dependent Variable

Since the inception of WPRS, benefit exhaustion has been the focal point of the identification component. P.L. 103-152 requires states to "identify which claimants will be likely to exhaust regular compensation..." Statistically, this suggests a binary outcome (i.e., only two possibilities); a claimant either exhausted regular unemployment insurance compensation or (s)he did not exhaust. Thus, the dependent variable in the DOL model was coded as "1" for exhaustees and as "0" for non-exhaustees. The output of the model is a predicted probability between zero and one that each claimant will exhaust benefits. Both the national and Maryland versions of the DOL model used logistic regression, the preferred statistical technique that accounts for the complexities introduced by a binary dependent variable.⁴ The advantages of logistic regression were also illustrated during each of the three DOL-sponsored Profiling Methods Seminars led by Dr. Robert St. Louis and held during the past year in Scottsdale, Arizona.

States using the same specification for the dependent variable in their WPRS models have typically used data elements reflecting the amount each claimant was paid over a complete benefit year to discern exhaustion. Two frequently-used definitions are: claimants with an ending balance of zero, or claimants paid amounts equal to or in excess of the total amount of UI benefits for which they were eligible. As mentioned, a binary dependent variable is a special, constrained case which usually cannot be modelled using simple ordinary least squares (OLS) regression analysis; a method must be used that accounts for the constraint. Of those that do, logistic regression best balances computational simplicity with theoretical and empirical reasoning.

(i) Alternative Specifications of the Dependent Variable

Since WPRS is an operational system, its practical effects must be considered along with its theoretical justification. In this context, some questions have been raised regarding the utility of a binary dependent variable. A few states correctly pointed out that this approach discards information; a claimant who almost exhausted is not distinguished from a claimant who came nowhere near exhausting, although the near-exhaustee may experience a greater need for reemployment assistance. Also, since benefits in most states are subject to variable potential duration, targeting likely exhaustees may result in some claimants with very low potential duration among those referred to re-employment services. As a result, some states have experimented with alternatives to a binary dependent variable representing exhaustion of regular unemployment compensation. These are discussed below:

Number of weeks claimed has been tested as a dependent variable using ordinary least squares (OLS) regression. This allows for distinctions between "near-exhaustees" and claimants who

⁴ See "Applied Logistic Regression" by David W. Hosmer, Jr. and Stanley Lemeshow (1989) for an in-depth treatment of the topic.

draw only a few weeks of benefits. However, constructing duration models is complicated by the dependent variable, which although continuous, is normally censored at 26 weeks.

The ratio of benefits drawn to benefit entitlement has been tested for the same reason as the number of weeks claimed, also using OLS. In one state, experimentation with this dependent variable concluded that using it in a WPRS model incurred significantly more estimation difficulties and gained little with respect to predictive capability. Ultimately, this method was abandoned in favor of logistic regression using a binary dependent variable.

***EVALUATION:** In theory, it is true that using a binary dependent variable ignores the distinction between near-exhaustees and claimants who collected only a few weeks of benefits. The utility of continuous dependent variables is predicated on the need to include near-exhaustees among the group assigned high probability values and referred to reemployment services. However, in both the aforementioned instances a censored sample is imposed and therefore questions of bias in estimation are raised. The meaning of the term "censored sample" is that since there is a maximum value for the dependent variable (i.e., 26 weeks, maximum benefit available), claimants who "exhaust" their benefits may still be unemployed and might draw more benefits if they were available to them. Having a maximum value on the dependent variable "censors" possible outcomes from exceeding whatever value has been set. Thus, it is typically impossible to obtain true outcomes in instances where claimants would claim (if allowed) a greater number of weeks than the benefit-week restriction. It therefore becomes necessary to apply a maximum likelihood estimation or a two-step procedure to provide unbiased parameter estimates when the continuous dependent variable is censored. (For more information on this topic, see Judge et al. (1985), pp. 780-785.) In general, since logistic regression is more straightforward and well-supported in economic literature, and since it focuses on the characteristics of claimants who exhaust benefits, it is the preferred method of targeting claimants for WPRS.*

(ii) Adjusting the Coding Scheme of the Exhaustion Variable

Developing a logistic regression model with a binary dependent variable still leaves open a number of options for experimentation. Some states have found that in certain instances, altering the coding scheme of the dependent variable proves useful. It is important to note that the coding scheme for the dependent variable affects the entire structure and function of the model; characteristics prevalent among the claimants coded as "exhaustees" will yield high predicted probabilities for current claimants who share those same characteristics, and vice-versa. In the DOL model, claimants are coded as exhaustees if they drew 100 percent of their entitlement and are coded as non-exhaustees if they did not draw 100 percent. Some states (see Table 2, beginning page 24) have found this definition of exhaustion too restrictive for their specific needs, and therefore have varied the definition of "exhaustion" in the following ways:

Expanding the scope of the exhaustion variable by using a more general definition is yet another method of separating the characteristics of "near-exhaustees" from those of other non-

exhaustees. For example, if at least 90 percent of benefits were depleted, claimant was coded as an exhaustee. This variation would cause the characteristics of both exhaustees **and** near-exhaustees to yield high probability scores for current claimants with the same characteristics. A related variation is to code claimants who exhaust a high percentage of benefits within a given time frame as exhaustees (e.g., 80 percent within 6 months of their benefit year begin (BYB) date). This would also expand the definition to include both exhaustees and near-exhaustees, and would also shorten the lag time for discerning exhaustion outcomes. Finally, exhaustion has also been redefined to automatically include claimants collecting EUC, since they had, by definition, exhausted regular benefits.

Narrowing the scope of the exhaustion variable by using a more restricted definition prevents the characteristics of certain exhaustees, who may not be considered in need of re-employment assistance, from yielding high probability scores. For example, some states have determined that claimants who take a full calendar year to exhaust 26 weeks of benefits are not truly in need of re-employment services; they may simply be collecting UI benefits between intervening spells of employment. To compensate, a time limit has been set (e.g., 8 months from BYB date) after which historic claimants would not be coded as exhaustees.

Weeks of potential duration has also been used as a criterion for narrowing the scope of the dependent variable. Variable duration complicates the use of exhaustion as the focal point of a model, because, for example, a claimant eligible for only 13 weeks of benefits has a higher probability of exhausting than a claimant eligible for a full 26 weeks of benefits, other things equal; yet, the 13 week claimant may not be determined to be in-need of reemployment assistance. To compensate, some states have chosen to set a minimum potential duration below which historical claimants cannot be coded as exhaustees. This is not a screen for current claimants and serves only to narrow the historical definition of exhaustion to claimants who actually collected UI for a significant length of time. On the other hand, some states have found that variable duration is not an issue because initial screens that exclude job-attached or seasonal claimants tend also to exclude those with low potential duration.

***EVALUATION:** Whether any of these techniques will be useful in a given state is a judgement best left to those familiar with that state's labor market trends. Expanding the definition of exhaustion would be most useful for states with low exhaustion rates because with only a small number of exhaustees, it is difficult to find characteristics that are widespread only among this scant few. By expanding the definition of exhaustion somewhat, more trends may become evident, making the model more reliable while still focusing on the long-term unemployed. Including claimants who drew at least 90 percent of their entitlement proved effective for at least one state with a low exhaustion rate. It is important to note, though, that a specific "cut-off" rate may be viewed as an arbitrarily selected point, and an appropriate "cut-off" level cannot remain fixed throughout time. An Appropriate level must rise during recessionary periods and fall during economic expansions. A careful evaluation of data may reveal some helpful trends, and lend support to the definition of the dependent variable (expanded or narrowed).*

Narrowing the definition of exhaustion using potential duration has been most useful for states that find that many short-duration--and perhaps seasonal--exhaustees pass all of the initial screens (e.g., recall, union hiring hall) yet are not truly in need of reemployment services. It ensures that the model focuses on exhaustees who are also long-term unemployed. Neither of the other narrowing criteria mentioned above (consecutive weeks, shortened time frame) have been tested conclusively as yet. However, it is important to note that such a technique would be predicated on the survival rate--the rate at which claimants continue to collect UI on a week-to-week basis. It is necessary to examine the survival rate at the specified cut-off point, whether it be six months or some other; a reliable relationship between the selected criteria and actual exhaustion must be established within this framework in order for such an approach to be tenable.

C. Independent Variables

(i) DOL Model Core Variables

While a few alternative definitions of the dependent variable have been tested, most experimentation has involved the independent variables. In the DOL model, five data elements used to develop a set of independent variables were suggested to states developing their WPRS systems. Some states adopted only these five elements and created state-specific versions of the DOL model such as the Maryland model. Others used the five elements as a starting point for analyzing a wider range of data. This section first summarizes each of the five key variables--education, job tenure, industry, occupation, and local unemployment rate--as they were implemented in the DOL model and then reviews and evaluates the findings of states surveyed concerning the use of the same elements.

Education is measured with a series of binary indicator variables which indicate that an inverse relationship exists between years of education and exhaustion. In the test state project, this specification found education to be a reliable predictor of exhaustion. The break points for the series of binary variables were developed partially by inferences from economic theory regarding impacts of education levels and partly by evaluating the historical data with which the model was developed.

Years of education squared was not included in the Maryland model but has been used by at least two other states to capture the marginal impact of education on exhaustion. This variable assumes that the relationship between education and exhaustion is not strictly linear, and therefore the quadratic representation of education is used in conjunction with a variable depicting the education levels linearly.

EVALUATION: *In most states, the same strong inverse relationship between education and exhaustion found in the DOL model was prevalent as well. However, there were a few notable exceptions where education was not a strong predictor. In at least two states, only the presence of a college degree had any significant impact on exhaustion, negative in both cases.*

Conversely, in another state, education was significantly correlated with exhaustion, but claimants with a college degree had the second-highest exhaustion probabilities (only those with less than a high school diploma were higher). Compared to possible alternatives (e.g., a continuous variable denoting years of education), the method of using binary indicators to model education is most appropriate. It emphasizes the importance of particular milestones--such as the attainment of a diploma or a degree--as opposed to individual years of schooling, which may have only marginal effects. A continuous variable simplistically assumes a constant linear relation, presumably negative, between years of education and exhaustion. However, including a quadratic term(s) along with a continuous variable relaxes the assumption of linearity and thus allows greater flexibility in determining education's impact. This may prove helpful to states that have had difficulty incorporating education thus far, although it will likely not contribute much to the overall predictive power of the model.

It is worth emphasizing that structural shifts in the labor market may necessitate re-examination of educational impacts, since different classes of workers may experience "dislocation" as factors such as technology, trade, and military downsizing keep the domestic economy in flux. Also, the relationship between education and exhaustion is should be viewed as sensitive to both the types of industries that drive primary local labor markets and to the demographic composition of the workforce. In areas where skill levels and educational backgrounds are fairly homogenous, it follows that education will not be a very effective predictor of exhaustion.

Job Tenure was used in the national model as series of binary indicators and in the Maryland model as a continuous metric. In retrospect, the continuous specification may in some ways overstate tenure's impact on exhaustion. This is because it assumes a constant impact--positive in the case of Maryland--over the distribution of values, meaning, for example, that the increase in exhaustion probability between 2 and 3 years of tenure is equal to the increase between 39 and 40 years. This is intuitively unlikely; a one-year change should exert more of a relative impact in the former case than in the latter. Thus, although a positive relationship exists, the unconstrained continuous variable may somewhat distort this relationship by assuming it is applicable to all values. A further concern is the integrity of tenure data, which can be suspect since claimants may have multiple base period employers or may have worked in one or more interim positions since being separated from their "real" occupation.

Years of job tenure squared is used in the same fashion as quadratic term for education described previously. The theoretical assumptions associated with this specification suggest that the relation between tenure and exhaustion is not strictly linear and therefore inclusion of the quadratic variable is necessary to accurately capture tenure's impact.

EVALUATION: *Several states have found that data on tenure are either unreliable or unavailable historically; therefore, tenure's utility for WPRS may not be fully realized for some time. Those with sufficient data have tested tenure's effects using several different specifications and many have obtained favorable results. Some states use a single binary variable set at a meaningful cut-off point, others use a series of binary variables representing several intervals, and still others use tenure as a continuous variable. With respect to this, one frequent difficulty*

with using tenure in a linear, continuous form is that in doing so, one assumes a constant marginal impact on exhaustion with each additional year of tenure. This assumption is often challenged by a graphical analysis of the relationship; an approach which has been found to satisfactorily express the truest relationship between tenure and exhaustion is to include a quadratic expression of tenure. While this is a better empirical specification of the relationship, it is more difficult to explain in practical application. An alternative method to capture the impact of tenure in selection and referral is that of "capping" the tenure variable by assigning a maximum value (e.g., for all observations 20 years and over). As these differences suggest, both the strength and direction of tenure's impact on exhaustion cannot be generalized across states and frequently vary within states as well. From this standpoint, including tenure squared (or some other non-linear form) may be productive if analysis suggests a non-linear relationship; tenure undoubtedly measures job-specific effects that are worth incorporating into profiling methodologies, but the challenge is in correctly identifying these effects in a model. Plotting the relationship between tenure and the dependent variable and using the results as a basis for creating and testing different variable specifications is the best way to approach this problem.

Industry was captured in the Maryland model using the Standard Industrial Classification (SIC) code(s) denoting a claimant's base period employer(s). Where multiple employers exist, the code corresponding to the separating employer was used. Some other states have used criteria such as earnings or tenure to discern the "primary" employer where necessary. However, it should be noted that no matter which employer it reflects, the SIC code by itself is not a meaningful variable and must be somehow transformed. In the test-state project, the SIC codes were aggregated to the industry division level and used to develop industry employment change rates. In very small industries, the change rates were weighted to reflect a more accurate impact on the labor market.

EVALUATION: *Since either industry or occupation must be used under the WPRS system and capturing occupational effects is difficult (see next section), most states have included industry in some form. Like Maryland, some have done this by attaching either historic or projected employment change rates to the code. Employment changes are typically calculated from the ES-202, Current Employment Statistics (CES), or similar data sources. This approach has proven effective for both models and screens in a number of states. However, shortcomings such as data lags have rendered growth rates ineffective in others. As alternatives, some states have either attached historic UI exhaustion rates to SIC codes or simply created a series of categorical variables from the code without attaching any additional information. Regardless of the form in which industry is depicted, almost all states have partially collapsed the SIC codes from the four-digit level in which they are typically recorded. This is because cell size at the four-digit level is typically too small to reflect the labor market a claimant faces. Most states have modelled industry variations at the division level or two-digit level, either statewide or within sub-state groupings. Given the fact that industry is widely available using a universal coding scheme, it is worthwhile for states to make every effort to include it meaningfully in their WPRS models.*

Occupation effects may be one factor that prevents industry from being a more powerful predictor. At the aggregate industry levels needed to achieve sufficient cell size, a wide range of skills and occupations exist within each. Measuring the relative demand for these occupations would undoubtedly aid the targeting of likely exhaustees. Unfortunately, occupational coding is a significant obstacle to both measuring such demand and to incorporating its effects into an operational system.

In the Maryland model, occupation was treated at the one-digit level and included as a series of binary variables. This had the effect of increasing the predicted probabilities of claimants in the relatively low-wage and high-exhaustion "clerical/sales" and "service" occupation groups.

***EVALUATION:** The specific occupational coding problems states have encountered are too numerous to mention here. In general, most involve either incomplete data or multiple coding schemes. In many states not all UI claimants are assigned an occupational code, creating a problem of missing data. Also, claimants may be assigned codes using one coding scheme (typically DOT--Dictionary of Occupational Titles), while data on historic or projected growth rates are organized using another scheme (typically OES--Occupational Employment Statistics). Although a "crosswalk" between coding schemes may be used, the added layer of complexity lessens the precision of the data because of conflicts in definitions, etc. Finally, the assignment of multiple codes (e.g., most recent occupation, desired occupation, etc.) and the complexity of the coding schemes makes the reliability of assigned codes an almost universal concern. Few states at this point have been able to incorporate meaningful occupational effects into their WPRS systems. Since occupation would seem to have a great deal of intuitive value in forecasting long-term unemployment, the challenge for the future is in developing reliable methods for coding claimants' occupations and collecting data that accurately measure the relative labor-market demand for them.*

Unemployment rate/sub-state variation refers to the unemployment rates and/or categorical variables used to control for regional variations in UI exhaustion. Even the smallest states exhibit a great deal of regional diversity; thus it should not be surprising that regional indicators are usually strong predictors of exhaustion. The Maryland model used the unemployment rate associated with each service delivery area (SDA).

***EVALUATION:** Most states that include unemployment rates in their models use data from the Local Area Unemployment Statistics (LAUS) program. Most often, recent measures of local unemployment rates are entered directly into the model; at least one state has experimented with additional trend measurements (e.g., percent change in unemployment rate). In states where unemployment and exhaustion are not as closely correlated, categorical variables are used as regional controls and/or as criteria for developing sub-state models. Regardless of the specific format of sub-state indicators, their primary function is as control variables; they do not normally aid in selecting likely exhaustees within a local office. This is because typically, a large majority of claimants in a given local office are from the same region and face the same labor market. Thus, sub-state indicators are usually significant predictors that serve to separate region-specific effects on exhaustion from those of variables (e.g., personal characteristics) that*

are more useful in selecting between individual claimants within local offices. Further discussion of this topic is included in section C-(iv), "Developing Sub-State Models."

(ii) Data Elements Beyond the DOL Model

While some states have used only the above five data elements and tailored them to their particular data and operations, others have used them as a starting point for more in-depth analyses. Such development and testing of additional variables is encouraged, provided either industry or occupation is included and all discriminatory variables are excluded. Several states have done a considerable amount of research, yielding the additional data elements listed in this section. This is a partial list, reflective only of the particular states included in this study and does not contain full details regarding specific data sources, transformations, etc. Further information on these processes may be obtained by contacting the UIS technical assistance team at the National Office.

Weekly benefit amount (WBA) has been experimented with in a variety of ways, and is often used in transformations of some other independent variables described below. WBA has also been used as a continuous variable, censored at the maximum amount, that captures the relationship between a claimant's benefit entitlement and his/her probability of exhaustion. In some form or another, the weekly benefit amount was used by eight of the thirteen States whose experiences are reflected in this paper.

***EVALUATION:** This variable is consistently a building block for strong predictors across many states and regions, but has been used on its own as well. Using WBA alone in a model discards information since no distinctions can be made between claimants eligible for the maximum weekly entitlement. Nonetheless, a number of states have found a positive and significant correlation between WBA and exhaustion using both continuous and categorical variables. Despite the variety of ways WBA is being used, it seems its most meaningful expression is as part of a wage replacement ratio, in conjunction with a control for potential duration. (See discussion below.)*

Wage replacement rate, the ratio of WBA to weekly base period wage, has generally been an effective data element for states that have tested it. Variables denoting wage replacement gain theoretical relevance by capturing the financial hardship involved in remaining unemployed and using UI benefits as a replacement for earnings. The larger the ratio, the less hardship exists for a claimant remaining unemployed; therefore this variable typically has a positive coefficient.

***EVALUATION:** Using the wage replacement rate has efficiently identified potential exhaustees in several states regardless of dominant industries or employment climates. This suggests that the replacement rate actually may capture a personal characteristic: it defines the "hardship" endured by remaining unemployed. The smaller the gap (a ratio value near one) between the weekly benefit amount and the weekly base period wage, the less of a fiscal incentive exists for a*

claimant to actively participate in a job search. However, at least one state found that although it accurately identifies exhaustees, it identifies primarily those with low potential duration who tend to have worked less during the base period and thus have a lower average weekly wage. This underscores the notion that, just because a variable is statistically significant, it does not necessarily follow that the variable is well-suited for inclusion in a WPRS system. Practical effects must be equally considered. With respect to this finding, it is logical to include a duration control in the model when using this variable or to test the WBA and/or wage variables separately in the model.

Base year wage is used to proxy two income-related factors: job skill level and reservation wage. Job skills are difficult to measure, given claim-taking constraints, but as a labor market measures employee value through salary, a higher wage is likely to be associated with higher skills. The reservation wage proxied through this variable identifies the minimum wage required for a claimant to accept work.

***EVALUATION:** Given the relevance of the aforementioned income-related factors that base year wage proxies, it has been used successfully as a building block for the wage replacement rate and as both a continuous and categorical variable on its own. One state that included base year wage as a continuous variable deflated its coefficient by the ratio of current average annual earnings to average annual earnings during the sample period. This technique--a variation of which was also applied to WBA in the same model--controls for the rate of inflation and ensures that current claimants' probabilities will not be artificially high(low) because of an accelerating(decelerating) rate of inflation relative to the sample period. Another variation used by at least two other states is to include the natural logarithm of the wage to compensate for an income distribution that is intuitively right-skewed by claimants with extremely high earnings.*

Potential duration of benefits has been used to control for claimants whose short duration of eligibility for UI benefits has essentially ensured exhaustion of their benefits. Claimants who have very short benefit duration have less time to complete their job search before their benefits run out and may be classified as exhaustees despite the fact that their personal characteristics may not be typical of the "dislocated worker" that WPRS is intended to serve.

***EVALUATION:** The relevance of controlling for potential duration depends on whether or not short-duration exhaustees are deemed in need of re-employment services and whether short-duration claimants tend to pass the initial screens for recall, hiring hall, etc. To the extent both of these are major issues in a state it may be necessary to control for potential duration. In using such a control, a state agency is implicitly defining their ideal group to be served. Therefore the duration issue needs to be evaluated from both a statistical and a policy perspective.*

The "separation" and "claim filed" dates have been used to develop a variable measuring the delay in filing for unemployment compensation. The delay is usually depicted as continuous in days or as a series binary indicators built from the continuous variable. The theory behind this

variable is that claimants who do not expect to have re-employment difficulty may not immediately file for UI benefits. Then, when they are unable to find suitable employment and turn to UI as a source of relief, they are in need of assistance. This variable has been found extremely significant with a positive effect in many of the states that have tested it. Four of the thirteen states were impressed enough with preliminary results to include it in their current model.

***EVALUATION:** While most states that have tested this variable discovered significant, positive effects, in at least one state it did not provide any appreciable predictive gains. In reviewing additional results, the delay variable appears to most effectively predict exhaustees in relatively urban labor markets. This is logical from standpoint that workers who start their unemployment spells with the expectation they will find suitable work, but cannot readily place themselves end up in particular need of job search assistance (JSA) in today's highly competitive job market. In rural areas, the relationship between filing delay and exhaustion is not as strong, perhaps because workers' skill sets tend to be more transferable and because a delay in filing may be more reflective of difficulty accessing a UI field office than of a choice to execute a job search independent of UI benefits and JSA. While the mostly positive results yielded by the filing delay variable make it a good variable with which to experiment, it is worth noting these potential limitations.*

Ratio of high quarter wage to base year wage controls for claimants whose base year earnings were accumulated primarily in one quarter. The larger the ratio, the less time spent working and earning wages during the base period. This variable has been found significant with a strong positive effect.

***EVALUATION:** If wage data are accessible, this is a worthwhile element to explore, since it is fairly easy to derive and seems to be applicable across a variety of labor markets. This ratio may capture wage replacement effects, since claimants with high ratios would not be accustomed to long-term earnings. It may also include intermittent workers with base period wages sufficient to qualify for UI. Finally, the ratio could reflect a lack of desirable personal characteristics such as employability and motivation and thus increase the probability of exhaustion.*

Number of base period employers controls for claimants who worked consistently during the base period, but for multiple employers. This element has been used as a binary variable indicating claimants with more than one base period employer, and as a continuous variable indicating the number of base period employers. Generally, it has shown a negative correlation between multiple employers and exhaustion probability. Five of our thirteen states have included this variable in their current model specification.

***EVALUATION:** There are many reasons for the statistical significance of this variable; one likely impetus is that claimants with multiple employers during a base period would have been between jobs at some point during their base period, making them familiar with the current*

dynamics of the job search process. These claimants may also have been intermittent workers or may have found a short-term job after their initial dislocation. However, it is also important to note that while these claimants may have the job search experience to aid themselves in finding a job, they are not necessarily placing themselves in positions they are able or willing to maintain in the long run. Tracking of the base period employers is useful given its explanatory power, but should be used with caution as a result of its tendency to rank at the bottom of the list claimants without a demonstrated capacity to maintain a long-term job.

Categorical representation of the month benefits began has been implemented with the intention of capturing the seasonality inherent in the month a claim is filed. Using a categorical variable representing each month of the year suggests that claimants filing in different months have different characteristics contributing to their probability of exhaustion.

***EVALUATION:** In states where monthly seasonality is not dramatic enough to be statistically significant, similar variables have been created which use quarterly identifiers to record seasonality. The propriety of this variable is to be considered with respect to the intended treatment of seasonal workers. Assuming that seasonal workers do not meet the definition of the "dislocated worker," use of a seasonality control is effective and useful. When a seasonal indicator is used as a variable in a statistical model, it leaves open the possibility that seasonal claimants could still end up being selected for referral to re-employment services. States in which this possibility presents a problem could consider using seasonal criteria as an initial screen rather than using an indicator variable. Another alternative might be to estimate the model using this variable to capture its effects, and then simply zero the associated coefficient when using the model to profile live claimants.*

(iii) Addressing Sub-state Labor Markets

In some states, dominant labor markets complicate the task of developing a reliable statewide model. For example, claimants living in urban areas, or working in large industries may exhaust benefits at different rates and in radically different patterns than claimants in the rest of the state. A statewide model that does not make some provision for such factors may be driven primarily by the dominant labor market. A model that identifies all of the claimants in urban areas as likely exhaustees simply because they come from high unemployment areas does nothing to identify exhaustion patterns in the rural parts of the state. The next two sections explain how to deal with dominant sub-state labor markets, both by controlling for them within a statewide model and by developing separate sub-state models.

Controlling for dominant labor markets: Controlling for dominant labor markets using binary variables creates a coefficient in the model for claimants from each labor market in question. If they exhaust at a higher (lower) rate than other claimants, the coefficient will be positive (negative). This helps to remove omitted variable bias that may otherwise have been exerted on the model's remaining coefficients and makes the model's predictions more reliable.

***EVALUATION:** It is appropriate to identify dominant labor markets, as explained above, but markets should be selected with theory and practice in mind. It is necessary to exhibit caution when developing a model as not to over-model the data; selection and identification of dominant labor markets should be both theoretically and statistically significant. The intent is simply to single out particular industries, occupations, or areas that, based on experience, are well-known to exhibit very different patterns and levels of exhaustion that cannot be explained by any of the other variables in the model.*

(iv) Developing Sub-state Models

When labor markets are vastly independent of one another and uniquely driven, some states have found that simple binary controls may still not allow them to target exhaustees as accurately as possible. When such structural change characterizes the labor markets within a state, sub-state models can be used to ensure that the independent variables' effects on exhaustion are measured as precisely as possible. A statistical F-test or Chi-square test can be used to test for structural change within a statewide data set.⁵ At least two sub-state modelling approaches have been successfully implemented thus far: regional models and industry models.

Regional models have been used where geography is considered the source of structural change within a state. For example, states that are primarily rural with one or two urban centers, large states, and states with several region-specific industries may be well served by regional models. An important caveat exists against using separate models for small, contiguous regions where considerable cross-commuting takes place. In this instance, otherwise similar claimants filing in the same local office can be profiled by different models and could receive sharply different probability scores based only on small differences in their area of residence. This is because separate data sets are used to develop the respective regional models and as such, they operate on different scales. The predicted probability values may not be comparable across models, meaning that claimants from different regions (and therefore profiled by different models) could not have their scores logically compared. Considerable overlap within local offices suggests that perhaps the regional boundaries are too narrow and may either need to be widened or expanded to the state level. States that have chosen regional models have typically created between four and fifteen models, each representing a logically defined group of counties or parishes (e.g., SDAs, MSAs).

Industry models involve the same logic as regional models, but have been used by states where the impacts of the independent variables on exhaustion are judged to vary more by industry than by geography. The key industries may not necessarily be regionally based, or other aspects of the state labor market may make regional models untenable. States that have chosen this approach have created between 10 and 15 models, primarily at the SIC division level, perhaps with sub-models for a few large two-digit groups (e.g., within the manufacturing division). Within

⁵ see Pindyck and Rubinfeld (1991).

division-level models, additional industry-based variation can still be incorporated at the two- or three-digit level using binary variables, employment change rates, or exhaustion rates.

EVALUATION: It is important to re-enforce the concept that labor markets should be examined for structural differences, changes or temporary shifts. For example, if job tenure were positively correlated with exhaustion in one region and negatively correlated in another, its value would be diminished in a state-level model. Or perhaps education only exerts influence on the exhaustion outcomes of workers in the manufacturing sector. In such cases, including an unemployment rate or a binary indicator as a control would account for different levels or rates of exhaustion, but would not account for the structural differences in tenure's or education's impact on exhaustion. Whether the differences lie in regional or industrial markets, it is important that the degree of structural difference is examined carefully and balanced against the practical impacts of using different models to assign probability scores.

III. POLICY IMPLICATIONS OF PROFILING METHODS

Choosing an appropriate methodology is a key factor in the successful implementation of WPRS. Since the statistical model or characteristic screens will largely control which claimants are targeted for re-employment services, policy issues inevitably will arise even in areas that may seem strictly technical.

The first issue is the need to **update models**. Models represent the historical period on which they were estimated. While initially useful and relevant to the process, an exceedingly old model becomes increasingly unrealistic and less useful over time. The second main issue is the necessity to balance a useful model with the **constraint of parsimony**. At some point in the model development, a maximum marginal benefit is reached. At that point, any additional variables added to the models "cost" more, in effort or computation, than the benefit they add to the overall process. The third issue is the **interpretation** of the model results. The meaning of the probability scores assigned by the statistical model tends to be a major focal point of the model output. The model provides a ranking mechanism by which claimants are selected for referral. It is important to remember that these scores are only **relative rankings**, and do not represent an absolute probability of exhaustion that could be used to compare claimants in different states. In other words, a .60 ranking in one state is not equivalent to a .60 ranking in all others. This issue has arisen with increasing frequency as the economy has improved and exhaustion rates have fallen, perhaps leaving available resources for re-employment services unallocated. If a claimant with a ranking of .49 is the highest on the probability list in a local office during a week, then that claimant has been identified by the methodology as "most in need of services", regardless of score. This scenario has led to inquiries as to whether there exists a score below which claimants, because their exhaustion probability is so low, would no longer benefit from re-employment services and should not be required to attend. A recent determination at the National Office level is that state selection and use of such a "threshold" is permissible, subject to Regional Office agreement that it has been implemented in a logical and productive fashion. Therefore, it is acceptable to use a threshold mechanism to prevent system flooding or referral of claimants who would no longer benefit from required services.

Another issue, encountered mostly by states using models with a small number of categorical variables, is **probability clustering**. Clustering occurs mainly when there are a small number of possible combinations among the independent variables in a statistical model, and therefore an equally small number of possible probability scores that could be assigned by the model. In this situation, it is important to have a mechanism in place that will randomly select the appropriate number of claimants to meet the service capacity guideline. A random selection mechanism is equally important when using a characteristic screening process, as the final selection pool will not be ranked in order of need for services; claimants are only identified as having passed the screening criteria. In both instances, the presence of a random selection mechanism is important from a legal standpoint. A common and simple random selection mechanism in place in several states uses the last four digits of the social security number. It should be added that with a statistical model, rather than just settling for probability clustering

and random selection, the clustering can be alleviated by adding or re-specifying independent variables such that the number of possible combinations increases. Provided this is not done haphazardly, it will produce a stronger and more reliable model.

Through the different stages of the implementation process, the TAT worked most intensively with the model developers in State offices. While common themes are likely to run throughout the system as a whole, some themes common to the model development segment are worth highlighting: first and foremost, WPRS is best viewed as a tool for both **identification and allocation**. It identifies those workers most in need of re-employment services and allocates the available supply of services accordingly. With respect to this, profiling, from the standpoint of identifying measurable factors that are accurate in predicting UI exhaustion, presents a difficult task. The methods with which we must work--whether characteristic screens or a statistical model--are imperfect ones, constrained by a number of empirical and political factors. However, both methods provide for more accurate forecasting of potential benefit exhaustees than is possible with less rigorous methods. Statistical modelling, because it weighs several factors simultaneously, is the most accurate identification method.

Secondly, it is also imperative to note that WPRS is much more than a theoretical forecasting exercise. It is a **practical application** of a system designed to identify, serve and track claimants on an ongoing basis. The system needs to be viewed as a whole by those working each part. Since the identification portion essentially drives the system, considerable forethought should be given to how it will affect the other parts of the system in an operational setting. For example, variables used to identify claimants as "likely to exhaust" must be legal and easily accessible, not just statistically significant. The benefits gained from the profiling approach should be commensurate with its data collection and automation costs; a trade-off exists between additional predictive ability and operational simplicity which generally favors a simple approach rather than an overly complex one. Finally, the group of claimants who tend to be identified as "likely to exhaust" should--assuming that benefit exhaustion is an accurate outcome measure--be consistent with the goals of WPRS. In short, profiling models should not be developed based solely on theoretical and statistical considerations. In fact, from a broad, system-wide perspective, the greatest value of a model is generally not found in any cryptic statistic, but rather in its application as a flexible allocation tool for matching the flow of claimants likely to exhaust with the available supply of re-employment services.

Lastly, the process of model development is a **dynamic one**. Currently, those claimants whose characteristics suggest they have the highest probabilities of exhausting UI are the first referred to re-employment services. Presumably, these services will reduce their likelihood of exhaustion such that, in the future, the same characteristics may not be found correlated with exhaustion. The estimation of profiling equations will need to evolve over time to avoid the omitted variable bias that could be otherwise introduced by the impact of re-employment services on exhaustion outcomes. This is likely to require controls for both the receipt of re-employment services receipt and for the types of services completed. Thus, the focus of profiling-related research is

likely to shift, and future DOL-sponsored Profiling Methods Seminars will address these relevant issues.

IV. CONTINUING CONCERNS

Maintaining Dynamic Processes and the Need for Persistent Evaluation and Redesign

Often when a program is brought about through an institution of a law, its implementation becomes a question of meeting legal requirements and satisfying terms agreed upon by parties involved. These circumstances seem to encourage those involved to set their goals to satisfying the basic criteria. As a result, the program then becomes fixed at the point where it has demonstrated legality. Although this level of development meets conformity requirements, it frequently falls short of the goals set by “the spirit of the law”. The explicit language of the law, written to address as many facets as possible, cannot possibly speak to future dynamics of the environment in which it must function.

On the heels of the 90's recession, Public Law 103-152 recognized a need to identify and refer dislocated workers to job search assistance services. Since that time the American economy has been growing and changing dramatically. Some partners in the WPRS System have interpreted this growth and change as a sign that the mechanism, designed to efficiently allocate resources in difficult times, was no longer a priority. When in fact, the changes our economy has experienced emphasize the need for maintenance of our systems as a whole. The composition of industries, the types of skills and the kinds of workers have all changed over the years. In contrast, the structure in which the WPRS must function is one that is reluctant to evolve. This reticence is likely a combination of the complexity of the inter-related systems and the fiscal constraints within which partners must work. In the past, difficulties surrounding bureaucratic evolution were sufficient to curb any urges to do so. However, the late 90's are very different times; we are facing a “...new economy [which] is increasingly driven by creativity, innovation and technology...”⁶ These forces promote change at breakneck pace and systems that remain inflexible will become obsolete. Those who are tasked with managing the WPRS system must accept this change and recognize its impact. The evolution of the job market should logically incite a reciprocal evolution of the worker profiling and Reemployment Service System, and the Unemployment Insurance System as a whole.⁷ Therefore, recognition of the changes in the labor market is a key to capturing information about how the system should evolve to be effective in a dynamic environment. In an economy experiencing growth, those who have difficulty finding work must truly be those who are in dire need of job search assistance. Building flexibility into the systems that serve these people will allow for optimal service provision to them, and therefore succeed at the task set forth by Public Law 103-152. Attention to maintaining a system which functions in a dynamic environment cannot be overemphasized. Varying points along the

⁶President Clinton, Comments on Unemployment Report and Job Training, March 6th, 1998

⁷For more information about the evolution of the Unemployment Insurance System, see “unemployment Insurance: A Dialogue”

business cycle will bring different volumes of claimants with different types of needs. It is imperative that the Profiling model and its service provision system be ready for the next phase of the cycle; this readiness is simply achieved by updating the selection criteria (updating models & coefficients) to reflect current economic environments and aligning available services with demands for environmentally current services and programs.

It may be clear that evaluation and redesign are in order, but often, multi-partner systems tend to fall prey to the problem that no one is willing to take ownership and/or responsibility for their part of the system as a whole. Once all partners have committed to the importance of the renovation and their roles in the process, the changes to the System can be managed effectively. Without complete agreement and cooperation from all those involved, the System risks becoming more fragmented and less efficient than its previous form. This underscores the importance of each partner's role in the system adjustment. Information collection, data modeling, referral processes, and services provided all must interface with the labor community in its evolved state. Each segment of an interactive system has an impact on, and receives input from, all other parts. One module that has not been updated can undermine the efficiency and effectiveness of the system as a whole.

In every process there are obstacles. Many of the issues that impair a system's ability to function at its maximum potential are surmountable. Things that are perceived as obstacles are often merely trails that have never been blazed. In an age of technological innovation and creativity, our systems have the potential to rise to the occasion and provide the population with an approach to job search assistance that exceeds all past expectations. Finding solutions to these extremely difficult tasks will be especially rewarding, particularly to those who have truly embraced their role in the process. The best strategies to develop plans for improvement are cooperative ones; not only inter-agency cooperation, but collaborative efforts across states as well can help provide the perspectives necessary to achieve the goal of the WPRS System. It seems clear from experiences in the last five years that WPRS cannot be a static system; it must ebb and flow with the tides of the labor force. As a result, issues at hand will change over time and new questions will arise continuously. Again, this reinforces the need for flexibility in modern systems. As individual systems are modified, States at different points of development will benefit from continued exchange of information. In addition to the phone/fax/on-site technical assistance that continues to be available through the TAT, three major vehicles for information dispersion are profiling methods seminars, the Information technology support center (ITSC) web site and additional research exchange documents. A number of states have completed formal documentation of their profiling methodologies and may be willing to have their product published. Development of a complete database of all states' methodologies has been frequently requested by some State partners. This is a suggestion that is currently under review at the National Office.

Through the variety of experiences encountered by the TA team, one main point remains abundantly clear: no single approach can best reflect the dynamics of all states. Each state's labor market is unique; so too are data and operational environments across states. State-specific testing and experimentation are the keys to building a model that is effective at distinguishing

exhaustees from non-exhaustees. Lessons learned from other states can serve as effective guides for research, but not as effective substitutes. The table that concludes this document summarizes the statistical modelling methodologies of 13 states with which the UIS TAT has had recent contact.

LESSONS LEARNED REFERENCE TABLE 2

	AK	WA	CT	ID	MI	MO	NV
Alternative dependent variable		90% exhaust, with minimum 20 wks potential duration		Number of weeks claimed			
Weekly benefit amount	Ratio with weekly base wage	Continuous linear form		Levels in linear form		Binary indicator for WBA > \$144	Levels in linear form
Base wage	Used in wage replacement ratio	Used in wage replacement ratio				Ratio with WBA, grouped into quartiles	
Benefit begin date	Categorical variables for quarter filed	Categorical variables for quarter filed	Binary Variable	Categorical variables indicating month filed			
Potential benefit duration		Linear continuous form		Categorical groupings			Number of quarters worked in last seven
Time between work end and claim filed dates		Continuous in days				Binary indicator for > 46 days	
Quadratic forms			Tenure				
Sub-state labor market classification	Categorical variables for local offices	Three sub-state models			Classification for occupation type (people, things,data) SDA identifiers		

	AK	WA	CT	ID	MI	MO	NV
Growth/ Decline Indicator	Industry		% change at industry levels; occupational codes	Industry exhaustion rates Growth and decline rates by industry			Vector of annual industry changes at 3 digit level
Number of base period employers	Average number of employers per quarter while working			Continuous number of employers		Binary indicator for more than one base period employer	Ratio of quarters worked for one employer over total quarters worked
Interactions			Tenure with education LMA with tenure LMA (or MSA) with industry LMA (or MSA) with industry concentration Industry concentration with industry CT TUR with Town TUR Town TUR w/ change in TUR				
Previous Claims	# of claims filed and exhausted within two years of current claim (continuous var. for exhausts, dummy for 0 exhausts)						
Ratio of high quarter wage over total base period wages					Grouped in quartiles		

	NJ	CA	SC	SD	TN	WV
Alternative dependent variable		90% exhaustion in 6 months after 1 st payment				
Weekly benefit amount			Linear continuous		WBA/BPW grouped into quartiles, represented with categorical variables	Range divided into quartiles + MBA
Base wage					See above	
Benefit begin date			Categorical variable representing year filed Categorical variable representing month filed			
Potential benefit duration	Used as control variable only		Number of weeks OR binary variable where weeks > 17			
Time between work end and claim filed dates			Linear, continuous representation in days		Categorical variables grouped by # of days delay	
Quadratic forms						

	NJ	CA	SC	SD	TN	WV
Sub-state labor market classification		Categorical variables representing 12 regional areas	Categorical variables representing counties			Separate industry models, exhaustion rates at two-digit level
Growth/Decline Indicator	% change in industry over two year period for LMA at division level				Statewide (at two digit level) binary variable for growth <= 2.5%	
Number of base period employers					Binary indicator for more than one base period employer	
Interactions		Education level * industry division	Duration >17 weeks* wage replacement rate	County unemployment rate * local office CUR* industry change CUR* occupation		Education * job Tenure * WBA range levels
Previous Claims						
Ratio of high quarter wage over total base period wages						

REFERENCES

The Worker Profiling and Reemployment Services System: Legislation, Implementation Process and Research Findings. Washington, DC: Unemployment Insurance Occasional Paper Series # 94-4, 1994.

Carey, James W. (1989). *Communications as Culture*. Boston: Unwin Hyman.

Hosmer, David W., Stanley Lemeshow. Applied Logistic Regression. New York, NY: John Wiley & Sons, 1989.

Kogan, Deborah, Kristin Wolff, Hugh Davies, Carole McCarthy, Cheryl Martin. (1997). "Framework for Collaboration: Partnering to Improve Welfare-to-Work Outcomes." Draft Paper

Judge, George G., et al. The Theory and Practice of Econometrics. New York, NY: John Wiley & Sons, 1985.

Pindyck, Robert S., Daniel L. Rubinfeld. Econometric Models & Economic Forecasts. New York, NY: McGraw-Hill, Inc., 1991.